

# Incentivized Campaigning in Social Networks

Bhushan Kotnis, Albert Sunny, Joy Kuri

**Abstract**—Campaigners, advertisers and activists are increasingly turning to social recommendation mechanisms, provided by social media, for promoting their products, services, brands and even ideas. However, many times, such social network based campaigns perform poorly in practice because the intensity of the recommendations drastically reduces beyond a few hops from the source. A natural strategy for maintaining the intensity is to provide incentives. In this paper, we address the problem of minimizing the cost incurred by the campaigner for incentivizing a fraction of individuals in the social network, while ensuring that the campaign message reaches a given expected fraction of individuals. We also address the dual problem of maximizing the campaign penetration for a resource constrained campaigner. To help us understand and solve the above mentioned problems, we use percolation theory to formally state them as optimization problems. These problems are not amenable to traditional approaches because of a fixed point equation that needs to be solved numerically. However, we use results from reliability theory to establish some key properties of the fixed point, which in turn enables us to solve these problems using algorithms that are linearithmic in maximum node degree. Furthermore, we evaluate the efficacy of the analytical solution by performing simulations on real world networks.

**Index Terms**—Social Networks, Incentivized Campaigns, Information Control, Bootstrapped Percolation, Reliability Theory

## I. INTRODUCTION

On-line social networking services have enabled advertisers, campaigners and activists to reach millions of individuals. In particular, the ability to recommend or share web articles [1], videos [2], and other media can be harnessed by campaigners for disseminating information to a worldwide audience. While such social network based campaigns sound promising, due to the fact that ideas propagate only a few hops from their origins [3], often, they are found to perform poorly in practice. Providing incentives for recommendations is a natural solution for increasing the hop count. For example, Dropbox, which offers cloud storage services, increased its customer base rapidly by offering incentives for social recommendations [4]. Although such referrals may increase the short term costs to the company by generating a large number of registrations for the free service, in the long term, the free registrations pay off, since a significant portion of free users eventually migrate to the premium service. Also, to keep the cost down, they capped the referral payouts — additional 500 MB only for the first 28 referrals, i.e. the referral payout is capped at 14 GB.

An alternative mechanism to limiting the cost is to cap the number of incentivized individuals, i.e., instead of capping in-

centive payments, offer large referral rewards<sup>1</sup> to a preselected set of individuals<sup>2</sup>. This approach allows the campaigner to target individuals who are best suited to use and recommend the service to their co-workers or friends.

### A. Related Work

In [5–7], the authors consider nodes to be either active or inactive. Making an inactive node active — not through network effects, but through direct intervention for kick-starting the campaign — is termed as “seeding.” Assuming full knowledge of the network structure, i.e., the adjacency matrix, the question asked in [5–7] is; Given a constraint on the number of seeds, what is the optimal set of seed nodes that would maximize the reach of the campaign? However, in several real-world settings, we may just have access to the statistical properties of the network. Also, the authors in [5–7] do not consider the cost incurred due to incentivization. Incentivization is the process of providing incentives to nodes who are already active, to persuade their friends to sign up. Incentivization happens throughout the campaign, whereas seeding, as discussed in [5], happens only at the beginning.

A related problem involving the calculation of an optimal pricing strategy for products sold to individuals in a social network was explored in [8, 9]. The authors in [8, 9] consider the problem of pricing a product and calculating the amount of cash-back (discount) that is provided to individuals as an incentive to evangelize the product. In this paper, we do not focus on optimal pricing, but rather focus on the size and cost of the campaign where individuals *do not* incur a cost to register for the service. The problems studied in this paper are more appropriate in settings where the service is free, or is based on a freemium model. Furthermore, we only assume partial knowledge of the network and also incorporate constraints that ensure a given campaign penetration, whereas the algorithms proposed in [8, 9] assume full knowledge of the network structure and do not consider any constraint on the campaign size.

The problem of computing the optimal referral payment mechanisms that maximize profit was studied in [10], by modelling the referral process as a network game. The authors in [10] conclude that a combination of linear payment mechanism (linear in the number of referrals) and threshold payment mechanism (payment only when number of referral exceeds a threshold) approximates the optimal pricing scheme. In this

<sup>1</sup>For example the living social website gives 100% cash-back on a purchased product if the customer persuades three others to buy the (same) product. Another famous example of uncapped referrals is Uber, and also the taxi service Lyft.

<sup>2</sup>Instead of advertising the referral scheme, a randomized scheme can be advertised. For example, the advertisement could be “those who register for the service will get a chance to win discounts for referring their friends.”

paper, we focus on the set of nodes to be incentivized while assuming that a pricing scheme, which can be computed based on the results in [10], is provided by the campaigner. Similar problems involving the computation of referral rewards in real time, for maximizing the campaign spread, were studied using the theory of optimal control in [11–15].

### B. Our Contributions

We consider a scheme where preselected incentivized individuals are presented with a reward when they register for the service, encouraging them to spread the news about the service to their friends. The decision of whether to offer an incentive to an individual is precomputed based on the solution of an optimization problem. We use a variant of the linear threshold model [5] for modelling the campaign spreading process. For a given fraction of such incentivized individuals, we first compute the campaign size (expected fraction of registered individuals) using bootstrap percolation. We then use this quantity to formulate the following optimization problems: 1). minimize the cost for achieving a given expected fraction of registered individuals, and 2). maximize the expected fraction of registered individuals for a given cost budget. These optimization problems are not amenable to traditional approaches because of a fixed point equation that needs to be solved numerically. However, we use results from reliability theory to establish some key properties of the fixed point, which in turn enables us to solve these problems using algorithms that are linearithmic in the maximum node degree. Through extensive simulations, we also study the efficacy our incentivization scheme in real world networks.

TABLE I  
A SUMMARY OF KEY NOTATION

Notation	Description
$p(k)$	fraction of degree $k$ nodes
$k_{max}$	maximum node degree of graph $\mathcal{G}$
$p_{th}(m k)$	probability of a degree $k$ node having threshold $m$
$\alpha_1$	activation probability of a type 1 node
$\alpha_2$	activation probability of a type 2 node
$\bar{d}$	mean node degree of graph $\mathcal{G}$
$q$	probability of encountering a type 2 node by traversing a randomly chosen link
$X_{k_2}$	random variables that denote the number of type 2 <i>active neighbours</i> of a degree $k$ node, given that there are $k_2$ type 2 neighbours
$Y_{k-k_2}$	random variables that denote the number of type 1 <i>active neighbours</i> of a degree $k$ node, given that there are $k - k_2$ type 1 neighbours
$u$	probability of reaching a registered node by following an arbitrary edge of the graph
$u_q$	fixed point of Equation (3)
$s(q)$	expected fraction of registered nodes at the end of the campaign
$s_k(q)$	expected fraction of degree $k$ registered nodes at the end of the campaign
$c_k$	cost of incentivizing a degree $k$ node
$\phi(k)$	probability of incentivizing a degree $k$ node
$\gamma$	minimum expected fraction of registered nodes
$\bar{c}$	maximum expected incentivization cost

The key notation used in this paper is summarized in Table I. The remainder of this paper is organized as follows. In Section II, we present the campaign model. In Section III, using

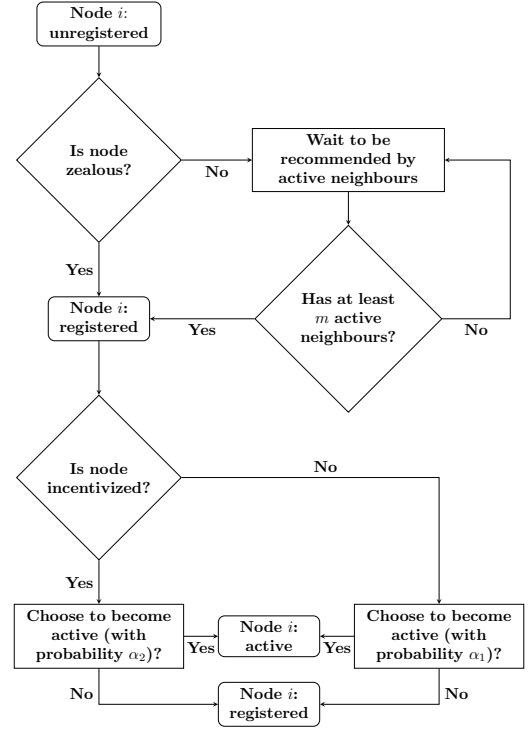


Fig. 1. Flow chart denoting the various stages of a node;  $m$  denotes the threshold of node  $i$ ,  $\alpha_1$  and  $\alpha_2$  are the activation probabilities of *incentivized* and *non-incentivized* nodes, respectively.

results from percolation theory, we compute the campaign size (expected proportion of registered individuals). Using this quantity, we formulate and solve two relevant optimization problems in Sections IV and V. In Section VI, we compare the analytical results with simulation performed on real-world networks. Finally, in Section VII, we conclude the paper.

## II. MODEL

Consider a set of  $\mathcal{N} = \{1, 2, \dots, n\}$  individuals who are connected to one another through a social network. For analytical tractability, as in [16, 17], we consider this network to be arbitrary, connected<sup>3</sup>, locally tree-like and uncorrelated<sup>4</sup>. We represent this network as an undirected graph  $\mathcal{G}(\mathcal{N}, \mathcal{L})$ , where  $\mathcal{N}$  and  $\mathcal{L}$  represent the set of nodes and links of graph  $\mathcal{G}$ , respectively. An undirected edge  $(a, b) \in \mathcal{L}$  if individuals  $a$  and  $b$  are neighbours in the underlying social network. In most scenarios, full knowledge of the network structure may not be available to the campaigner. In such cases, the campaigner can obtain statistical properties of the network through data mining. One such property is the degree distribution. Let  $\{p(k), k \geq 1\}$  be the degree distribution of graph  $\mathcal{G}$ .

We consider a campaign on the network represented by graph  $\mathcal{G}$  (see Fig. 1 for the flow chart). The nodes in graph  $\mathcal{G}$  can be either in “*active*”, “*registered*” or “*unregistered*” state. Once active, a node remains active. Active nodes are very spirited and express their strong support for the campaign

<sup>3</sup>The configuration model random graph is connected with high probability if and only if the minimum degree is greater than or equal to 3.

<sup>4</sup>Uncorrelated networks are networks where the degree of a node is statistically independent of the degree of any other node in the network [18].

by encouraging their neighbours to register, whereas nodes in the registered and unregistered states refrain from recommendations. For a node to become active, it must first show support for the campaign by registering itself. However, all registered nodes need not be active.

The campaign starts with a set of *zealous* individuals — those who intrinsically desire the service. These nodes do not need any recommendation to register, and are registered for the campaign from the very start. If a zealous node becomes active, it will recommend the service to its neighbours. On the other hand, nodes that are not zealous, register only if the number of active neighbours exceeds a predefined threshold value (see Fig. 1). Let  $p_{th}(m|k)$  denote the probability of a degree  $k$  node having threshold  $m$ . This is a generalization of the celebrated linear threshold model [5] — a model of choice for studying the dissemination and adoption of new products, technologies and ideas [19–22]. We note that, by definition,  $\sum_{k \geq 1} p(k) \cdot p_{th}(0|k)$  denotes the fraction of zealous nodes in the network<sup>5</sup> — fraction of nodes that register without any recommendations.

A node that receives a reward when it registers for the service is more likely to tell its neighbours about the reward and encourage them to register. Therefore, we divide the nodes into two categories: the ‘*non-incentivized*’ (type 1) and the ‘*incentivized*’ (type 2). The campaigner decides the fraction of incentivized nodes before the start of the campaign<sup>6</sup>. However, these nodes become aware of the fact that they are incentivized only if they register for the service<sup>7</sup>. Since incentivized nodes are provided with incentives upon registration, they are more likely to be enthused, and will play an active part in the campaign than non-incentivized nodes. Let  $\alpha_1 \in (0, 1]$  and  $\alpha_2 \in (0, 1]$  be the probabilities of a non-incentivized (type 1) and incentivized (type 2) node becoming active, respectively. Due to the presence of incentives, we assume that  $\alpha_2 > \alpha_1$ .

### III. COMPUTING THE CASCADE SIZE

Before presenting the problems, we need to first compute cascade size, i.e., the expected fraction of nodes that have registered at the end of the campaign. Let  $p(k'|k)$  be the probability of encountering a degree  $k'$  node by traversing a link from a degree  $k$  node. From [16], we know that this conditional distribution, for an uncorrelated network, has the following form  $p(k'|k) = \frac{k' \cdot p(k')}{\bar{d}}$ , where  $\bar{d}$  denotes the mean degree of the graph  $\mathcal{G}$ . Let  $q(k)$  be the probability of encountering a type 2 node by traversing a randomly chosen link from a degree  $k$  node. Then, we have

$$\begin{aligned} q(k) &= \sum_{k' \geq 1} \Pr[\text{type 2 neigh.} | k'] \cdot p(k'|k) \\ &= \sum_{k' \geq 1} \phi(k') \cdot p(k'|k) = \frac{1}{\bar{d}} \sum_{k' \geq 1} k' \phi(k') p(k') \end{aligned}$$

<sup>5</sup>Campaigner can estimate the fraction of zealous nodes by conducting surveys on a reasonable sized demographic.

<sup>6</sup>Alternatively, whenever an individual registers, the campaigner can dynamically decide whether to incentivize this individual. This problem is similar to stochastic control problems, and is beyond the scope of this paper.

<sup>7</sup>Incentives are assigned assuming knowledge of the exact degree of the registered node, because the exact node degree can be obtained after registering. For example, one can ask individuals to register through an on-line social network, and compute the exact degree from their contact list.

Since  $q(k)$  is independent of  $k$ , let  $q(k) = q \forall k \in \{1, 2, \dots, k_{max}\}$ . Here,  $k_{max}$  is the maximum node degree.

Since the network is locally tree-like, with some effort, it can be shown that the number of type 2 neighbours of a degree  $k$  node is a *binomial random variable* with the probability distribution function  $\hat{p}(k_2|k) = \binom{k}{k_2} \cdot q^{k_2} \cdot (1-q)^{k-k_2}$ . Let  $p_{ex}(k')$  be the excess degree distribution, i.e., the degree distribution of a node encountered by following a randomly chosen link without counting that link. For an uncorrelated network, from [17], we know that  $p_{ex}(k') = \frac{(k'+1) \cdot p(k'+1)}{\bar{d}}$ .

Now, let  $u$  denote the probability of reaching a registered node by following an arbitrary edge of the graph. First, let us consider an arbitrary node  $j$  of degree  $k$  and threshold  $m$ . Next, we compute the probability that node  $j$  registers due to recommendations from its active neighbours, given that we have arrived at node  $j$  by following an arbitrary link in the graph. For ease of presentation, we do not explicitly write the condition “following an arbitrary link.”

$$\begin{aligned} &P[j \text{ registers} | j \text{ is of degree } k, j \text{ has threshold } m] \\ &\stackrel{(a)}{=} \sum_{l=m}^k P[j \text{ has } l \text{ active neighbours} | j \text{ is of degree } k] \\ &= \sum_{l=m}^k \sum_{k_2=0}^k P[k_2 \text{ type 2 neighbours} | j \text{ is of degree } k] \\ &\quad \cdot P[l \text{ active neigh.} | j \text{ is of degree } k \text{ and } k_2 \text{ type 2 neigh.}] \\ &= \sum_{l=m}^k \left( \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot P[X_{k_2} + Y_{k-k_2} = l] \right) \\ &= \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot P[X_{k_2} + Y_{k-k_2} \geq m] \end{aligned} \quad (1)$$

where Equality (a) follows because the events  $\{j \text{ registers}\}$  and  $\{j \text{ has } m \text{ or more active neighbours given that } j \text{ has threshold } m\}$  are equivalent, and the events  $\{j \text{ has } l \text{ active neighbours}\}$  and  $\{j \text{ has threshold } m\}$  are independent of each other.

In Equality (1),  $X_{k_2}$  and  $Y_{k-k_2}$  are random variables that denote the number of type 2 and type 1 *active neighbours* of a degree  $k$  node, given that there are  $k_2$  type 2 and  $k - k_2$  type 1 neighbours. With some effort, it can be shown that random variables  $X_{k_2}$  and  $Y_{k-k_2}$  are independent and have a binomial distribution with parameters  $(k_2, \alpha_2 u)$  and  $(k - k_2, \alpha_1 u)$ , respectively. Here,  $\alpha_1$  and  $\alpha_2$  are the probability of a type 1 and type 2 node becoming active, respectively. The independence occurs because the probability of incentivizing a node is independent of the degree of its neighbouring nodes. The binomial distribution arises here because the network is locally tree-like.

From Equation (1), it is evident that the probability of a node registering is *independent of its type*<sup>8</sup>. Therefore, the probability that by following an arbitrary link we can reach a node that registers due to recommendations is given by

$$\begin{aligned} &P[\text{node registers due to recommendation}] = \sum_{k \geq 1} p_{ex}(k) \cdot \\ &P[\text{node reg. due to recommendations} | \text{excess degree } k] \end{aligned}$$

<sup>8</sup>A node registers either if it is zealous (this event is independent of its type), or through recommendations which depend on its neighbours and type of neighbours but not on its type.

$$\begin{aligned}
f(q, u) &= \sum_{k \geq 0} p_{ex}(k) \cdot p_{th}(0|k+1) + \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} p_{th}(m|k+1) \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot P[X_{k_2} + Y_{k-k_2} \geq m] \\
&= \sum_{k \geq 0} p_{ex}(k) \cdot p_{th}(0|k+1) + \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} p_{th}(m|k+1) \sum_{k_2=0}^k \binom{k}{k_2} \cdot q^{k_2} \cdot (1-q)^{k-k_2} \cdot \\
&\quad \left( \sum_{l+l' \geq m} \binom{k_2}{l} (\alpha_2 u)^l (1-\alpha_2 u)^{k_2-l} \binom{k-k_2}{l'} (\alpha_1 u)^{l'} (1-\alpha_1 u)^{k-k_2-l'} \right) \\
g_k(q, u) &= p_{th}(0|k) + \sum_{m \geq 1} p_{th}(m|k) \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot P[X_{k_2} + Y_{k-k_2} \geq m] \\
\text{where } X_{k_2} &\sim \text{Bin}(k_2, \alpha_2 u) \text{ and } Y_{k-k_2} \sim \text{Bin}(k-k_2, \alpha_1 u)
\end{aligned}$$


---

$$\begin{aligned}
&= \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} P[\text{node registers, node has threshold } m] \\
&\quad \text{node has excess degree } k] \\
&\stackrel{(a)}{=} \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} p_{th}(m|k+1) \cdot P[\text{node registers} \\
&\quad \text{node has excess degree } k, \text{ node has threshold } m] \\
&= \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} p_{th}(m|k+1) \cdot \\
&\quad \sum_{k_2=0}^k \binom{k}{k_2} \cdot q^{k_2} \cdot (1-q)^{k-k_2} \cdot P[X_{k_2} + Y_{k-k_2} \geq m] \quad (2)
\end{aligned}$$

On the RHS of Equation (2), we use the excess degree distribution because we discount the link that we followed to arrive at the node. In Equality (a), we use  $p_{th}(m|k+1)$  because if we include the link on which we arrived, a node of excess degree  $k$  will have  $k+1$  links. Arguing along the lines of [17], we can conclude that  $u$  has to satisfy the following self consistency equation

$$\begin{aligned}
u &= P[\text{node is zealous}] \\
&\quad + P[\text{node reg. due to recommendations}] \\
&= \sum_{k \geq 0} p_{ex}(k) p_{th}(0|k+1) \\
&\quad + P[\text{node reg. due to recommendations}] \\
&= f(q, u) \quad (3)
\end{aligned}$$

where  $f(q, u)$  is as given at the top of this page.

For any  $q$ , Equation (3) is a fixed point equation in  $u$ . However, due to the complex nature of the function  $f(q, u)$ , the existence of a  $u$  that satisfies Equation (3) is not obvious. Using results from reliability theory, in Proposition 3.2, we prove the existence and uniqueness of the fixed point of Equation (3). Before that, the following proposition establishes the nature of function  $f(q, u)$  (w.r.t  $u$ ).

**Proposition 3.1:** If the fraction of zealous nodes lies in the open interval  $(0, 1)$ , then  $\forall q \in [0, 1]$ ,  $f(q, u)$  is a continuously differentiable, convex, monotonically increasing function of  $u$ .

*Proof:* In Appendix B-A. ■

Proposition 3.1 gives us some intuition about the existence and uniqueness of the fixed point. However, we establish the same rigorously in the following proposition.

**Proposition 3.2:** For every  $q \in [0, 1]$ ,

- (i) If no node in the network is zealous, then  $u = 0$  is the only solution of Equation (3).

- (ii) If every node in the network is zealous, then  $u = 1$  is the only solution of Equation (3).
- (iii) If the fraction of zealous nodes lies in the interval  $(0, 1)$ , then Equation (3) has a unique fixed point in  $(0, 1)$ .

*Proof:* In Appendix B-B. ■

For ease of presentation, in the remainder of the paper, we assume that the fraction of zealous nodes lies in the open interval  $(0, 1)$ . For any  $q \in [0, 1]$ , let  $u_q$  denote the fixed point of Equation (3). Since the fixed point is unique, we can obtain  $u_q$  by iteratively solving Equation (3). Our next proposition qualitatively describes the variation of the fixed point  $u_q$  as function of  $q$ .

**Proposition 3.3:** If  $\alpha_2 > \alpha_1$ , then  $u_q$  is a strictly increasing continuous function of  $q$ , i.e.,  $\frac{\partial u_q}{\partial q} > 0$ .

*Proof:* In Appendix C. ■

Now, given a  $q \in [0, 1]$ , let  $s_k(q)$  be the expected fraction of degree  $k$  nodes that have registered at the end of the campaign. The expression for  $s_k(q)$  can be obtained as follows

$$\begin{aligned}
s_k(q) &= P[\text{node registers} | \text{degree } k] \\
&= P[\text{node is zealous} | \text{degree } k] \\
&\quad + P[\text{node reg. due to recommendations} | \text{degree } k] \\
&= \sum_{m \geq 1} P[\text{node reg, node has threshold } m] \\
&\quad \text{node has degree } k] + p_{th}(0|k) \\
&= p_{th}(0|k) + \sum_{m \geq 1} p_{th}(m|k) \cdot \\
&\quad P[\text{node reg} | \text{node has degree } k, \text{ node has threshold } m] \\
&= g_k(q, u)|_{u=u_q} \quad (4)
\end{aligned}$$

where  $g_k(q, u)$  is as given at the top of this page and  $u_q$  is the fixed point of Equation(3). Given a  $q \in [0, 1]$ , let  $s(q)$  be the expected fraction of nodes that have registered at the end of the campaign.  $s(q)$  is also termed as the size of the epidemic. Then, using arguments similar to the ones used to derive Equation (5), it can be shown that

$$s(q) = \sum_{k \geq 1} p(k) \cdot g_k(q, u)|_{u=u_q} = \sum_{k \geq 1} p(k) \cdot s_k(q) \quad (5)$$

Once again using results from reliability theory, in Proposition 3.4, we establish the increasing and non-decreasing nature of functions  $g(q, u)$  and  $g_k(q, u)$  (with respect to  $u$  and  $q$ ).

**Proposition 3.4:** If  $\alpha_2 > \alpha_1$ , then

- 1)  $\frac{\partial g(q, u)}{\partial q} > 0$  and  $\frac{\partial g(q, u)}{\partial u} > 0$
- 2)  $\frac{\partial g_k(q, u)}{\partial q} \geq 0$  and  $\frac{\partial g_k(q, u)}{\partial u} \geq 0$ .

*Proof:* Similar to that of Propositions 3.1 and C.1. ■

#### IV. COST MINIMIZATION UNDER CASCADE SIZE CONSTRAINT

Providing incentives is a costly affair. The campaigner may either be interested in minimizing the cost while guaranteeing that a given proportion of population registers, or in maximizing the registrations for a given cost budget. In this section, we look at the former problem.

##### A. Cost of incentivization

Incentives provided by the campaigner is usually a function of its degree because the number of potential recommendations depend on the degree. Let  $c_k$  be the cost of incentivizing a degree  $k$  node. *Incentivized nodes obtain incentives only if they registers.* Therefore, the expected cost per incentivized degree  $k$  node is given by  $c_k \cdot s_k(q)$ , where  $s_k(q)$  (see Equation 4) is the probability that an incentivized degree  $k$  node registers. Let  $\phi(k)$  denote the probability of incentivizing a degree  $k$  node. Then, the average cost per node for incentive-policy ( $\phi = \{\phi(k), k \geq 1\}$ ) is given by  $\sum_{k \geq 1} p(k) \phi(k) \times$  (cost of node of deg.  $k$ ) =  $\sum_{k \geq 1} p(k) \cdot c_k \cdot \phi(k) \cdot s_k(q)$ , where  $q = \frac{1}{d} \sum_{k \geq 1} k \cdot p(k) \cdot \phi(k)$ .

##### B. Problem formulation

Minimizing the cost while providing guarantees on the number of expected registrations is appropriate for campaigns where the campaigners are mandated to achieve a given target. This problem can be mathematically formulated as follows

$$\min_{0 \leq \phi \leq 1} \sum_{k \geq 1} c_k \cdot p(k) \cdot \phi(k) \cdot s_k(q) \quad (6)$$

$$\text{Subject to: } s(q) \geq \gamma \text{ and } q = \frac{1}{d} \sum_{k \geq 1} k \cdot p(k) \cdot \phi(k)$$

where  $\gamma \in [0, \gamma_{max}]$  is the minimum expected fraction of registered individuals that must be achieved and  $\bar{d}$  is the mean degree of the network. Here,  $\gamma_{max}$  is the expected fraction of registered individuals obtained by incentivizing everyone. The expression for  $s_k(q)$  (Equation (4)) and  $s(q)$  (Equation (5)) involve  $u_q$ , which is the solution to the fixed point Equation (3). Thus, it is not possible to apply traditional analytical techniques such as the *Karush-Kuhn-Tucker (KKT) conditions* to solve the above problem. Furthermore, the problem may be non convex, and applying numerical techniques such as *genetic algorithms*, or *Markov Chain Monte Carlo* methods may not yield a globally optimal solution.

##### C. Solution approach

In this section, we present transformations that allow us to efficiently compute the global optima of the cost minimization problem presented in the previous section. To help us with this, we state and prove the following proposition. The intuition behind this proposition is that, as  $\alpha_2 > \alpha_1$ , increasing  $q$  increases the proportion of type 2 nodes, which results in a higher  $s(q)$ .

**Proposition 4.1:** If  $\alpha_2 > \alpha_1$ , then function  $s : [0, 1] \rightarrow [\gamma_{min}, \gamma_{max}]$  is a monotonically increasing bijection, where

$\gamma_{max} = s(0)$  and  $\gamma_{max} = s(1)$  are the expected fraction of registered individuals obtained by incentivizing nobody and everybody, respectively.

*Proof:* We have

$$\begin{aligned} \frac{\partial s(q)}{\partial q} &= \frac{\partial (g(q, u)|_{u=u_q})}{\partial q} \\ &\stackrel{(a)}{=} \frac{\partial g(q, u)}{\partial q} \Big|_{u=u_q} + \frac{\partial g(q, u)}{\partial u} \Big|_{u=u_q} \cdot \frac{\partial u_q}{\partial q} \\ &= \underbrace{(> 0)}_{\text{Proposition 3.4}} + \underbrace{(> 0)}_{\text{Proposition 3.4}} \cdot \underbrace{(> 0)}_{\text{Proposition 3.3}} \\ &\Rightarrow \frac{\partial s(q)}{\partial q} > 0 \end{aligned} \quad (7)$$

where Equality (a) follows from an application of the chain rule. From Inequality (7), we can see that  $s(q)$  is a monotonically increasing function defined on the closed interval  $[0, 1]$ . Hence,  $s : [0, 1] \rightarrow [\gamma_{min}, \gamma_{max}]$  is a bijection. ■

Proposition 4.1 establishes that  $s(q)$  is a monotonically increasing function of  $q$  and is a bijection. Therefore, we can replace the constraint  $s(q) \geq \gamma$  with the constraint  $q \geq q_\gamma$ , where  $q_\gamma$  is such that  $s(q_\gamma) = \gamma$ . Since  $s(q)$  is a monotonically increasing function of  $q$ , we can compute  $q_\gamma$  by performing a line search over the closed interval  $[0, 1]$ . The following algorithm does this.

---

##### Algorithm 1 Algorithm to compute $q_\gamma$

---

```

1: Set  $q_1 := 0$  and  $q_2 := 1$ 
2: while  $|s(q_1) - s(q_2)| > \epsilon$  do
3:    $q_{mp} := (q_1 + q_2)/2$ 
4:   Solve equation  $u = f(q_{mp}, u)$  to obtain  $u_{q_{mp}}$ .
5:   if  $s(q_{mp}) < \gamma$  then
6:      $q_1 := q_{mp}$ 
7:   else
8:      $q_2 := q_{mp}$ 
9:   end if
10: end while
11: return  $q_\gamma := (q_1 + q_2)/2$ 

```

---

Proposition 4.1 does not tell us how the expected fraction of registered degree  $k$  nodes ( $s_k(q)$ ) responds to changes in  $q$ . While, the function  $s_k(q)$  may not be a bijection for all values of  $k$ , we show that it is a *non-decreasing* function of  $q$ , for all values of  $k$ .

**Proposition 4.2:** If  $\alpha_2 > \alpha_1$ , then for all  $k \geq 1$ ,  $\frac{\partial s_k(q)}{\partial q} \geq 0$

*Proof:* By first writing the total derivative in terms of the partial derivatives (see Proposition 4.1), and then by using the non-negativity of these partial derivatives (from Propositions 3.3 and 3.4). ■

Now, if  $\gamma \leq \gamma_{min} = s(0)$ , then the cascade constraint is met without incentivizing anyone. In such cases, the optimum incentive policy and the optimum cost is given by  $\{\phi(k) = 0, \forall k \geq 1\}$  and 0, respectively. On the other hand, if  $\gamma = \gamma_{max} = s(1)$ , then the cascade constraint is met only if everyone is incentivized, i.e.,  $\{\phi(k) = 1, \forall k \geq 1\}$ , and the optimum cost is given by  $\sum_{k \geq 1} c_k \cdot p(k) \cdot s_k(1)$ . For the cases when  $\gamma \in (\gamma_{min}, \gamma_{max})$ , due to Propositions 4.1, the

cost minimization problem can be re-written as follows

$$P_1 : \min_{0 \leq \phi \leq 1} \sum_{k \geq 1} c_k \cdot p(k) \cdot \phi(k) \cdot s_k(q) \quad (8)$$

$$\text{Subject to: } q \geq q_\gamma \text{ and } q = \frac{1}{\bar{d}} \cdot \sum_{k \geq 1} k \cdot p(k) \cdot \phi(k)$$

where  $q_\gamma$  is chosen such that  $s(q_\gamma) = \gamma$ . Since  $\gamma \in (\gamma_{\min}, \gamma_{\max})$ , we have  $q_\gamma \in (0, 1)$  (see Proposition 4.1). The next proposition establishes the existence of an optimal policy, of problem  $P_1$ , that satisfies all its the constraints, with equality.

**Proposition 4.3:** If  $\alpha_2 > \alpha_1$ , then problem  $P_1$  has an optimal solution  $\phi^{opt} = \{\phi_k^{opt}, k \geq 1\}$  such that  $\sum_{k \geq 1} k \cdot p(k) \cdot \phi^{opt}(k) = \bar{d} \cdot q_\gamma$

*Proof:* In Appendix D-A. ■

From Proposition 4.3, we can see that problem  $P_1$  has an optimal solution that satisfies all its constraints with equality. Therefore, to obtain a solution of problem  $P_1$ , we restrict our search to the set of policies that satisfy the constraints of problem  $P_1$  with equality. Subsequently, a change of variable gives us the following linear program

$$P_2 : \min_{\{\nu_k, k \geq 1\}} \bar{d} \cdot q_\gamma \cdot \sum_{k \geq 1} \nu_k \cdot \mu_k(q_\gamma) \quad (9)$$

$$\text{Subject to: } \sum_{k \geq 1} \nu_k = 1 \text{ and } 0 \leq \nu_k \leq \frac{k \cdot p(k)}{\bar{d} \cdot q_\gamma} \quad \forall k \geq 1$$

where  $\mu_k(q_\gamma) = \frac{s_k(q_\gamma) \cdot c_k}{k}$ . The following proposition establishes a crucial relation between the optimal solutions of problems  $P_1$  and  $P_2$ .

**Proposition 4.4:** Let  $\{\nu_k^{opt}, k \geq 1\}$  be an optimal solution of problem  $P_2$ . If  $\alpha_2 > \alpha_1$ , then we can obtain an optimal solution of problem  $P_1$  as follows

$$\phi^{opt}(k) = \begin{cases} 0 & \text{if } p(k) = 0 \\ \frac{\nu_k^{opt} \cdot \bar{d} \cdot q_\gamma}{k \cdot p(k)} & \text{otherwise} \end{cases}$$

*Proof:* In Appendix D-B. ■

Therefore, to obtain an optimal solution of problem  $P_1$ , we just need to solve  $P_2$  and apply Proposition 4.4. Problem  $P_2$  is a linear program and can be solved by any well-known LP-solver in polynomial time ( $O((k_{\max})^{3.5})$  time, where  $k_{\max}$  is the maximum node degree in the network)[23]. However, we exploit the rich structure of this problem and solve it in linearithmic time  $O(k_{\max} \cdot \log(k_{\max}))$

Let  $\mathcal{S} = \{1, 2, \dots, k_{\max}\}$ . Here,  $k_{\max}$  is the maximum degree of the network. Consider a permutation  $\sigma : \mathcal{S} \rightarrow \mathcal{S}$  such that if  $k_1 < k_2$ , then  $\mu_{\sigma(k_1)}(q_\gamma) \leq \mu_{\sigma(k_2)}(q_\gamma)$ , i.e.,  $\mu_{\sigma(1)}(q_\gamma) \leq \mu_{\sigma(2)}(q_\gamma) \leq \dots \leq \mu_{\sigma(k_{\max})}(q_\gamma)$ . The permutation  $\sigma$  can be obtained in  $O(k_{\max} \cdot \log(k_{\max}))$  time by sorting the set  $\{\mu_k(q_\gamma), k \geq 1\}$  in ascending order. Using this order, we can obtain an optimal solution  $\nu^a$  of problem  $P_2$  in  $\Theta(k_{\max})$  time by following the steps in Algorithm 2.

In Algorithm 2,  $\nu_k^a$  denotes the  $k^{th}$  component of vector  $\nu^a$ . We illustrate the principle behind Algorithm 2 using a jar filling analogy. Variable  $\nu_{\sigma(j)}^a$  can be interpreted as a jar with capacity  $\frac{\sigma(j) \cdot p(\sigma(j))}{\bar{d} \cdot q_\gamma}$ . The for loop in the algorithm is equivalent to filling these jars with water from another jar of unit capacity, in an ascending order, as given by the

**Algorithm 2** Algorithm to compute optimal solution of  $P_2$

---

```

1:  $res := 1$  and  $\nu^a := [0, 0, \dots, 0]$  — a  $1 \times k_{\max}$  vector
2: for  $j \in [1, 2, \dots, k_{\max}]$  do
3:   if  $res > 0$  then
4:      $\nu_{\sigma(j)}^a = \min \left\{ res, \frac{\sigma(j) \cdot p(\sigma(j))}{\bar{d} \cdot q_\gamma} \right\}$ 
5:      $res := res - \nu_{\sigma(j)}^a$ 
6:   end if
7: end for
8: return  $\nu^a$ 

```

---

permutation  $\sigma$ . From steps 1, and 4 – 5 of Algorithm 2, it is easy to see that  $\nu^a$  is a feasible solution of problem  $P_2$ . The optimality of  $\nu^a$  with respect to problem  $P_2$  is established by the following proposition.

**Proposition 4.5:** The vector  $\nu^a$  returned by Algorithm 2 is an optimal solution of problem  $P_2$ .

*Proof:* In Appendix E. ■

## V. CASCADE SIZE MAXIMIZATION UNDER BUDGET CONSTRAINT

In this section, we consider the challenge faced by a resource-constrained campaigner. The campaigner desires to maximize the campaign size, but is constrained by limited resources. Let  $\bar{c}$  denote the limit on the expected incentivization cost. Then, we can formulate the cost-constrained cascade maximization problem as follows

$$P_3 : \max_{0 \leq \phi \leq 1} s(q)$$

$$\text{Subject to: } \sum_{k \geq 1} c_k \cdot p(k) \phi(k) s_k(q) \leq \bar{c}$$

$$\text{and } q = \frac{1}{\bar{d}} \sum_{k \geq 1} k \cdot p(k) \cdot \phi(k)$$

### A. Solution approach

Let  $s^{opt}$  be the optimum value of problem  $P_3$ . From Proposition 4.1, we know that there exists a unique  $q^{opt} \in (0, 1)$  such that  $s(q^{opt}) = s^{opt}$ . From Proposition 4.1, we know that  $\frac{\partial s(q)}{\partial q} > 0$ , i.e.,  $s(q)$  is a monotonically increasing bijection. Therefore, it can be argued that  $q^{opt}$  can be obtained by solving the following optimization problem

$$P_4 : \max_{0 \leq \phi \leq 1} q$$

$$\text{Subject to: } \sum_{k \geq 1} c_k \cdot p(k) \phi(k) s_k(q) \leq \bar{c}$$

$$\text{and } q = \frac{1}{\bar{d}} \sum_{k \geq 1} k \cdot p(k) \cdot \phi(k)$$

Let us define  $\mu_k(q) = \frac{s_k(q) \cdot c_k}{k}$  and  $\nu_k(q) = \frac{k \cdot p(k) \cdot \phi(k)}{\bar{d} \cdot q}$ . Then, problem  $P_4$  can be re-written as follows

$$P_5 : \max_{\nu, q \in (0, 1)} q$$

$$\text{Subject to: } \bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k(q) \cdot \mu_k(q) \leq \bar{c}$$

$$\sum_{k \geq 1} \nu_k(q) = 1 \quad \text{and} \quad 0 \leq \nu_k(q) \leq \frac{k \cdot p(k)}{\bar{d} \cdot q} \quad \forall k \geq 1$$

Now, given a  $q \in (0, 1)$ , consider the following criteria

- (1)  $\bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k(q) \cdot \mu_k(q) \leq \bar{c}$
- (2)  $\sum_{k \geq 1} \nu_k(q) = 1$
- (3)  $0 \leq \nu_k(q) \leq \frac{k \cdot p(k)}{\bar{d} \cdot q} \quad \forall k \geq 1$

If we can find a vector  $\nu$  that satisfies the above criteria, then the 2-tuple  $(\nu, q)$  is a feasible solution of problem  $P_5$ . Let  $\nu^a(q)$  be the vector output by Algorithm 2 when  $q_\gamma$  is set to  $q$ . From Proposition 4.5, we know that this vector minimizes the LHS of criterion (1) subject to criteria (2) and (3). Therefore, if we have  $\bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k^a(q) \cdot \mu_k(q) > \bar{c}$ , then for the given value of  $q$ , no vector  $\nu$  can satisfy criterion (1). Now, let  $\nu^a(q + \Delta q)$  be the vector output by Algorithm 2 when  $q_\gamma$  is set to  $q + \Delta q$ . Then, for any  $\Delta q > 0$ , if  $\bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k^a(q) \cdot \mu_k(q) > \bar{c}$ , then we have

$$\begin{aligned} \bar{c} &< \bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k^a(q) \cdot \mu_k(q) \stackrel{(a)}{\leq} \bar{d} \cdot q \cdot \sum_{k \geq 1} \nu_k^a(q + \Delta q) \cdot \mu_k(q) \\ &\stackrel{(b)}{\leq} \bar{d} \cdot (q + \Delta q) \cdot \sum_{k \geq 1} \nu_k^a(q + \Delta q) \cdot \mu_k(q + \Delta q) \end{aligned} \quad (10)$$

where Inequality (a) follows because the vector  $\nu^a(q)$  is an optimal solution and  $\nu^a(q + \Delta q)$  is just a feasible point at  $q$ , and Inequality (b) follows because  $\frac{\partial \mu_k(q)}{\partial q} = \frac{c_k}{k} \cdot \frac{\partial s_k(q)}{\partial q} \geq 0$  (see Proposition 4.2). From Inequality 10 we can conclude that if there does not exist a vector  $\nu$  that satisfies criteria (1)-(3) for some  $q \in (0, 1)$ , then no vector satisfies these criteria for any  $q' \in (q, 1)$ . Therefore, the optimal value  $q^{opt}$  is given by the largest value of  $q$  in  $[0, 1]$  such that vector  $\nu^a(q)$  satisfies criteria (1), and can be computed using the following algorithm

---

**Algorithm 3** Algorithm to compute optimum of  $P_3$

---

```

1: Set  $q_1 := 0$  and  $q_2 := 1$ 
2: while  $|s(q_1) - s(q_2)| > \epsilon$  do
3:    $q_{mp} := (q_1 + q_2)/2$ 
4:   Set  $q_\gamma = q_{mp}$  in Algorithm 2 and obtain  $\nu^a(q_{mp})$ 
5:   if  $\bar{d} \cdot q_{mp} \cdot \sum_{k \geq 1} \nu_k^a(q_{mp}) \cdot \mu_k(q_{mp}) > \bar{c}$  then
6:      $q_2 := q_{mp}$ 
7:   else
8:      $q_1 := q_{mp}$ 
9:   end if
10: end while
11: Return  $q^{opt} := (q_1 + q_2)/2$ 

```

---

Algorithm 3 does a line search to find the largest values of  $q \in [0, 1]$  such that the tuple  $(\nu^a(q), q)$  meets criteria (1)-(3).

## VI. SIMULATION AND NUMERICAL EVALUATIONS

We formulated and solved the optimization problems assuming an uncorrelated, locally tree-like and connected network. However, real world social networks may contain triads, loops, multiple connected components, and need not be uncorrelated. Despite such differences, we show that, in most cases, our analytical results closely match the simulations on real world social networks. We evaluate our analytical results on two very different networks: Gnutella [24] — a p2p file sharing network, and Hamsterster [25] — a social network for people with pet hamsters. Table II presents a few statistical measures of these networks. The degree distributions of these networks

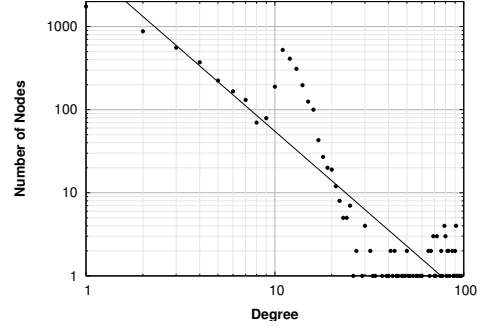


Fig. 2. Degree distribution of network A (Gnutella network); the thick black line denotes the best fit of the power law to the degree distribution.

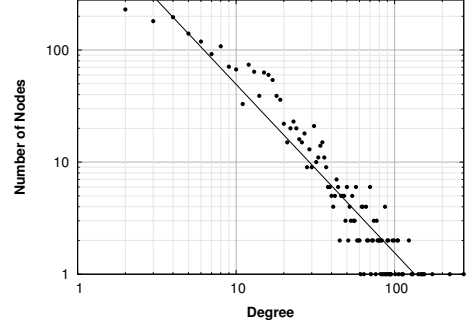


Fig. 3. Degree distribution of network B (Hamsterster network); the thick black line denotes the best fit of the power law to the degree distribution.

are displayed in Figures 2 and 3. All simulations are averaged over 10000 runs. To enhance legibility, error bars are suppressed in all the plots.

TABLE II  
SIMPLE PARAMETERS OF THE TWO REAL-WORLD NETWORKS USED FOR SIMULATIONS.

	Network A	Network B
Source	p2p-Gnutella08 [24]	Hamsterster [25]
Network category	Peer-to-peer	Social
Nodes	6301	2426
Edges	20777	16631
Maximum degree	273	97
Average degree	6.59	13.71
Number of triangles	2383	53265
Clustering coefficient	0.01	0.51
Connected components	2	148

We first evaluate the correctness of the analytical calculation of the campaign size, i.e.,  $s(q)$ . The simulation plots were generated by considering a linear threshold model where nodes register if at least 50% of their neighbours are active. We assume that 30% of the nodes in the network were zealous. For a different values of  $q$ , we studied three schemes:  $\phi^{all}$  — incentivizing all nodes equally with probability  $q$ ,  $\phi^{high}$  — incentivizing nodes starting from the highest degree till degree  $k'$  such that  $q = \frac{1}{\bar{d}} \sum_{k \geq k'} p(k) \phi^{high}(k)$ , and  $\phi^{low}$  — incentivizing nodes starting from the lowest degree till degree  $k'$  such that  $q = \frac{1}{\bar{d}} \sum_{1 \leq k \leq k'} p(k) \phi^{high}(k)$ . The analytical results were obtained by extracting the degree distribution from the networks.

The number of triangles in the p2p Gnutella network is very small, which is also reflected in the clustering coefficient. Therefore, the Gnutella network is very similar to a locally

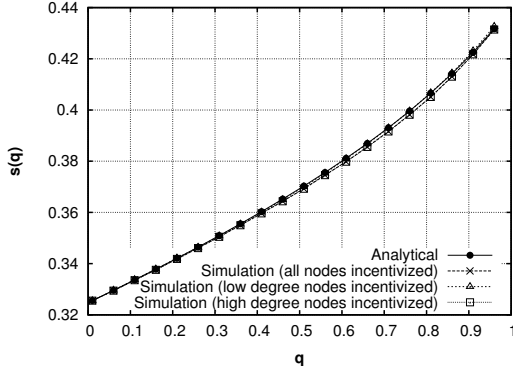


Fig. 4. Analytical and simulated values of  $s(q)$  on network A (Gnutella network).

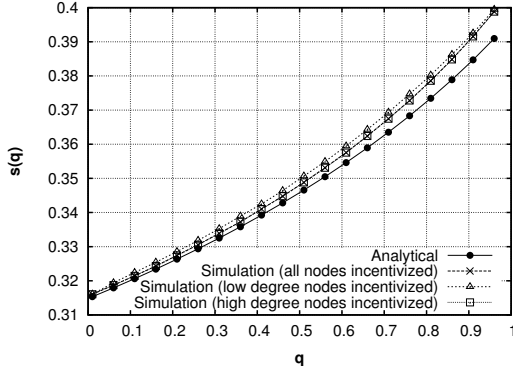


Fig. 5. Analytical and simulated values of  $s(q)$  on network B (Hamsterster network).

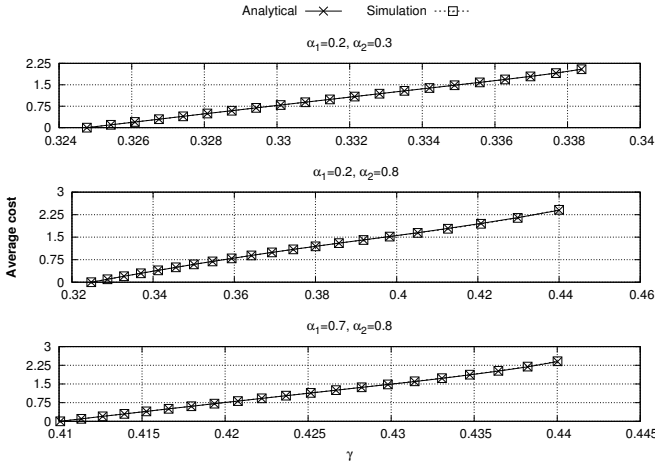


Fig. 6. Analytical and simulated values of average cost vs.  $\gamma$  for network A (Gnutella network).

tree-like network. As consequence of this, for the Gnutella network, the simulation and analytical results are in excellent agreement with each other (see Fig. 4). On the other hand, one can observe a deviation of the simulation results from the analytical on Hamsterster for large values of  $q$  (see Fig. 5). This behaviour may in part be due to the presence of significant number of triangles and loops in the network.

Next, we consider a linear incentive structure, where an incentivized node obtains a unit reward per neighbour, i.e.,  $c_k = k$ ,  $\forall k \geq 1$ . For the cost minimization problem, for a given  $\gamma$ , using the degree distributions, we analytically

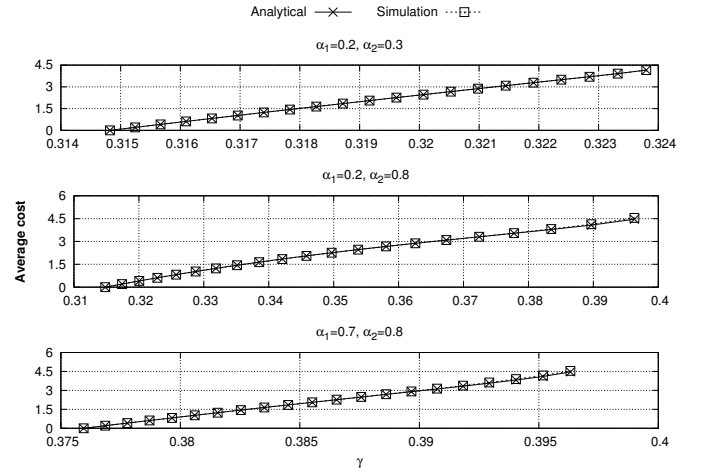


Fig. 7. Analytical and simulated values of average cost vs.  $\gamma$  for network B (Hamsterster network).

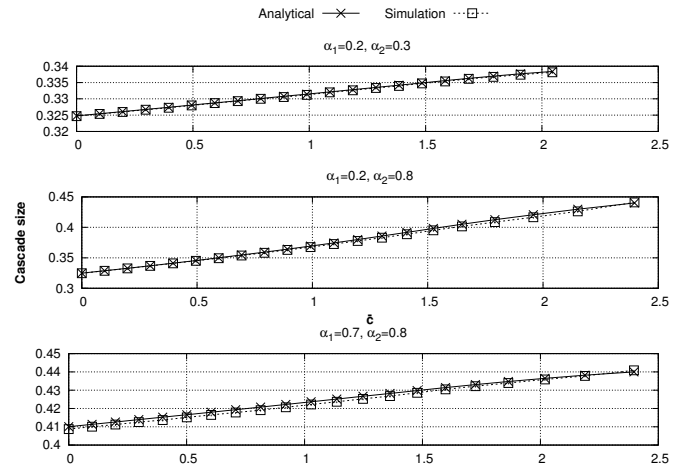


Fig. 8. Analytical and simulated values of cascade size vs.  $\bar{c}$  for network A (Gnutella network).

compute the solution  $\phi$  and the average cost. We then simulate the linear threshold process (nodes register if at least 50% of their neighbours are active) on the respective real world networks using the analytical solution  $\phi$ , and compare the average cost obtained from analysis and simulations (see Fig. 6 and Fig. 7). Similarly, for the campaign size maximization problem we analytically compute  $\phi$  and size  $s$  for a given cost budget  $\bar{c}$ . The analytical solution is then used in the simulation of cascade maximization problem (see Fig. 8 and Fig. 9).

As observed in Fig. 6 and 7, the analytical average cost is in excellent agreement with the simulated one for different values of  $\alpha_1, \alpha_2$  and  $\gamma$ , on both the networks. Similarly Fig. 8 and 9 show an excellent match between simulation and analytical computation of campaign size in almost all the plots. For large values of  $\alpha_1$  and  $\alpha_2$ , the simulated campaign size in the Hamsterster network is larger than the analytically computed campaign size. This may in part be due to the large number of triangles in the network. This suggests that in social networks containing large number of triangles and loops, the analytically computed size represents a lower bound.



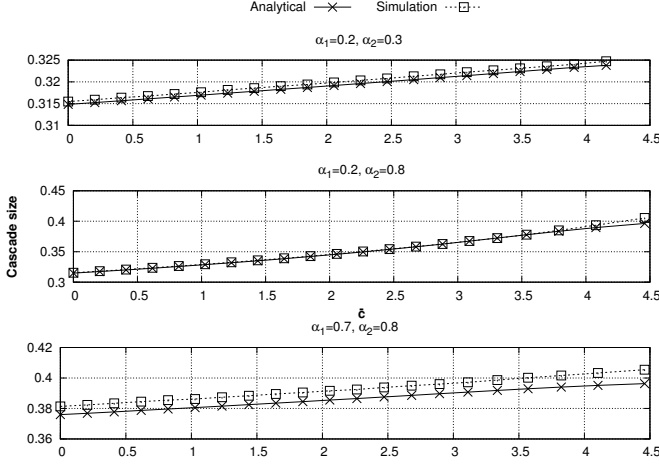


Fig. 9. Analytical and simulated values of cascade size vs.  $\bar{c}$  for network B (Hamsterster network).

## VII. CONCLUSION

In this paper, we have studied the problem of campaigning in social networks by offering incentives for referrals. We used ideas from percolation theory to compute the campaign size, which was then used to formulate two optimization problems. These problems were not amenable to traditional solutions since they involved a fixed point equation whose solution was analytically intractable. We used results from reliability theory to establish some key properties of the fixed point that enabled us to solve these problems with simple algorithms having linearithmic time complexity. Although we assumed an uncorrelated and locally tree-like network in the analysis, through extensive simulations on real world social networks, we showed that our analytical results are applicable in real world networks.

## APPENDIX A

### SOME RESULTS FROM RELIABILITY THEORY

In this paper, we prove several propositions using the theory of *stochastic order*. Let  $X$  and  $Y$  be two random variables. Then,  $X$  is said to be smaller than  $Y$  in the *usual stochastic order* (denoted by  $X \leq_{st} Y$ ) [26] if and only if

$$P[X > x] \leq P[Y > x] \quad \forall x \in (-\infty, \infty)$$

Next, we present two theorem from [26] without proof, and a lemma. We will use these theorems and lemma in several of our proofs.

**Theorem A.1:**  $X \leq_{st} Y$  if and only if  $\mathbb{E}[\psi(X)] \leq \mathbb{E}[\psi(Y)]$  holds for all non-decreasing functions  $\psi$  for which the expectation exists.

*Proof:* Refer Chapter 1 of [26]. ■

**Theorem A.2:** The usual stochastic order is closed under convolutions, i.e., If  $X \leq_{st} Y$  and  $X' \leq_{st} Y'$ , then  $X' + X \leq_{st} Y' + Y$ .

*Proof:* Refer Chapter 1 of [26]. ■

**Lemma A.1:** (i) Let  $I(p_1)$  and  $I(p_2)$  be two Bernoulli random variable with parameters  $p_1$  and  $p_2$ , then  $I(p_1) \leq_{st} I(p_2)$  if and only if  $p_1 \leq p_2$ .

(ii) Let  $Z(n, p_1)$  and  $Z(n, p_2)$  be two binomial random variable with parameters  $(n, p_1)$  and  $(n, p_2)$ , then for any  $n \in \mathbb{N}$ ,  $Z(n, p_1) \leq_{st} Z(n, p_2)$  if  $p_1 \leq p_2$ .

*Proof:* Part (i): By comparing the complementary cumulative distribution function of the two random variable, and by applying the definition of *usual stochastic order*.

Part (ii): A binomial random variable with parameter  $(n, p)$  is the sum of  $n$  i.i.d. Bernoulli random variable with parameter  $p$ . Now, the result follows from an application of Part (i), and due to the fact that stochastic order is closed under convolution (Theorem A.2). ■

We also use some results from the theory of *stochastic convexity* in our proofs. For the sake of completeness, we reproduce some definitions and theorems from [26]. Let  $\{X(\theta), \theta \in \Theta\}$  be a parametrized collection of random variables. We say

- (a)  $\{X(\theta), \theta \in \Theta\}$  is *stochastically increasing (SI)* if  $\mathbb{E}[\psi(X(\theta))]$  is non-decreasing in  $\theta$  for all non-decreasing functions  $\psi$ .
- (b)  $\{X(\theta), \theta \in \Theta\}$  is *stochastically increasing and convex (SICX)* if  $\{X(\theta), \theta \in \Theta\} \in SI$  and  $\mathbb{E}[\psi(X(\theta))]$  is non-decreasing and convex in  $\theta$  for all non-decreasing convex functions  $\psi$ .
- (c)  $\{X(\theta), \theta \in \Theta\}$  is *stochastically increasing and linear (SIL)* if  $\{X(\theta), \theta \in \Theta\} \in SICX$  and  $\mathbb{E}[\psi(X(\theta))]$  is non-decreasing and concave in  $\theta$  for all non-decreasing concave functions  $\psi$ .

Note that, by definition, we have  $\{X(\theta), \theta \in \Theta\} \in SIL \Rightarrow \{X(\theta), \theta \in \Theta\} \in SICX$ .

**Lemma A.2:** Let  $X(n, p)$  be a binomial random variable with parameters  $(n, p)$ . Then,  $\{X(n, p), p \in (0, 1)\} \in SICX$ .

*Proof:* By combining example Example 8.B.3 and Theorem 8.B.9 of [26], it can shown that  $\{X(n, p), p \in (0, 1)\} \in SIL$ . Then, due to the fact that  $SIL \Rightarrow SICX$ , the lemma follows. ■

**Theorem A.3:** Suppose  $\{X(\theta), \theta \in \Theta\}$  and  $\{Y(\theta), \theta \in \Theta\}$  are two collection of random variables such that  $X(\theta)$  and  $Y(\theta)$  are independent for each  $\theta$ . If  $\{X(\theta), \theta \in \Theta\} \in SICX$  and  $\{Y(\theta), \theta \in \Theta\} \in SICX$ , then  $\{X(\theta) + Y(\theta), \theta \in \Theta\} \in SICX$ .

*Proof:* Refer the proof of Theorem 8.A.15 in [26]. ■

**Theorem A.4:** Suppose that for each  $\theta \in \Theta$ , the support of  $X(\theta)$  is in  $\mathbb{N}$ . Then,  $\{X(\theta), \theta \in \Theta\} \in SICX$  if and only if  $\{X(\theta), \theta \in \Theta\} \in SI$  and  $P[X(\theta) > m]$  is non-decreasing and convex in  $\theta$  for all  $m \in \mathbb{N}$ .

*Proof:* Refer the proof of Theorem 8.A.10 in [26]. ■

## APPENDIX B

### EXISTENCE AND UNIQUENESS OF THE FIXED POINT

In this section, we show that for any  $q \in [0, 1]$ , the fixed point Equation 3 has a unique solution. The proof in this section is split into two parts. In the first part, we show that function  $f(q, u)$  is monotonically increasing and convex in  $u$ . In the section part, we use the result of the first part to establish the existence and uniqueness of the fixed point.

### A. Proof of Proposition 3.1

The differentiability of  $f(q, u)$  follows from its definition. In this section, we show that  $f(q, u)$  is monotonically increasing and convex in  $u \in (0, 1)$ . To show this, we use the theory of stochastic convexity.

Let  $\{X(k_2, \alpha_2 u), u \in (0, 1)\}$  and  $\{Y(k - k_2, \alpha_1 u), u \in (0, 1)\}$  be two collection of independent binomial random variables with parameters  $(k_2, \alpha_2 u)$  and  $(k - k_2, \alpha_1 u)$ , respectively.

Then, from Lemma A.2, we have  $\{X(k_2, \alpha_2 u), u \in (0, 1)\} \in SICX$  and  $\{Y(k - k_2, \alpha_1 u), u \in (0, 1)\} \in SICX$ . Since the random variables are independent, an application of Theorem A.3 gives us  $\{X(k_2, \alpha_2 u) + Y(k - k_2, \alpha_1 u), u \in (0, 1)\} \in SICX$ . Now, by applying Theorem A.4 we can conclude that

- (1)  $\{X(k_2, \alpha_2 u) + Y(k - k_2, \alpha_1 u), u \in (0, 1)\} \in SI$ .
- (2)  $P[X(k_2, \alpha_2 u) + Y(k - k_2, \alpha_1 u) > m]$  is a *non-decreasing convex* function of  $u \in (0, 1)$  for all  $m \in \mathbb{N}$ .

Let  $h(u, k, k_2, m) = P[X(k_2, \alpha_2 u) + Y(k - k_2, \alpha_1 u) > m]$ . We note that function  $h(u, k, k_2, m)$  is non-decreasing and convex in  $u$ . Since the network is connected and the fraction of zealous nodes lies in the open interval  $(0, 1)$ , there exists positive integers  $k_0, m_0$  such that  $m_0 \in (0, k_0]$  and  $p_{ex}(k_0)p_{th}(m_0|k_0 + 1) > 0$ . Therefore, we have

$$\begin{aligned} \frac{\partial f(q, u)}{\partial u} &= \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 1} p_{th}(m|k + 1) \cdot \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot \\ &\quad \frac{\partial h(u, k, k_2, m)}{\partial u} \cdot (1 - \mathbb{I}_{\{k=k_0, m=m_0, k_2=k\}}) \\ &\quad + p_{ex}(k_0)p_{th}(m_0|k_0 + 1) \frac{\partial P[X(k_0, u\alpha_2) \geq m_0]}{\partial u} \stackrel{(a)}{>} 0 \end{aligned}$$

where Inequality (a) follows because  $h(u, k, k_2, m)$  and  $P[X(k_0, u\alpha_2) \geq m_0], m_0 \in (0, k_0]$  are non-decreasing and monotonically increasing functions of  $u$ , respectively. Similarly, we have

$$\begin{aligned} \frac{\partial^2 f(q, u)}{\partial u^2} &= \sum_{k \geq 1} p_{ex}(k) \sum_{m \geq 0} p_{th}(m|k + 1) \sum_{k_2=0}^k \hat{p}(k_2|k) \cdot \\ &\quad \frac{\partial^2 h(u, k, k_2, m)}{\partial u^2} \stackrel{(b)}{\geq} 0 \end{aligned}$$

where Inequality (a) follows because  $h(u, k, k_2, m)$  is convex in  $u$ . ■

### B. Proof of Proposition 3.2

We recollect that  $u$  denotes the probability of finding a registered node by following an arbitrary link of the network. Nodes with zero threshold do not need any recommendations, and are registered from the start of the campaign. Thus, if every node has threshold value zero, then the entire network is registered and we have  $u = 1$ . On the other hand, nodes with non-zero threshold have to be recommended by active nodes to register. Therefore, if all nodes in the network have non-zero threshold, then there are no node from which the campaign can start. Hence, in such scenarios, we have  $u = 0$ .

Now, let us consider the case when the fraction of zealous nodes lies in the interval  $(0, 1)$ . Let us define  $f^0(u) =$

$f(q, u) - u$ . It is easy to see that  $u_0$  is a fixed point of Equation 3 if and only if  $f^0(u_0) = 0$ . By differentiating  $f^0(u)$  with respect to  $u$  twice, we get  $\frac{\partial^2 f^0(u)}{\partial u^2} = \frac{\partial^2 f(q, u)}{\partial u^2} \geq 0$ , i.e.,  $f^0$  is a convex function of  $u$ . We also have

$$\begin{aligned} f^0(0) &= f(q, 0) = \sum_{k \geq 0} p_{ex}(k) \cdot p_{th}(0|k + 1) \stackrel{(a)}{>} 0 \\ f^0(1) &= f(q, 1) - 1 \stackrel{(b)}{<} 0 \end{aligned}$$

where Inequality (a) follows because the network is connected and the fraction of zealous nodes lies in the open interval  $(0, 1)$ , and Inequality (b) follows because the  $f(q, 1)$  is a convex combination of non-negative terms less or equal to 1 (some terms are strictly less than 1). Since  $f^0(0) > 0$  and  $f^0(1) < 0$ , due to the continuity of function  $f^0$ , there exist a point  $u_0 \in (0, 1)$  such that  $f^0(u_0) = 0$  i.e.,  $f(q, u_0) = u_0$ .

We prove the uniqueness of the fixed point by contradiction. Now, if  $u_0, u_1 \in (0, 1)$  are two fixed points of Equation (3), then we should have  $f^0(u_0) = f^0(u_1) = 0$ . Without loss of generality, let us assume that  $0 < u_0 < u_1 < 1$ . Now, choose positive real numbers  $u_2 \in (u_1, 1)$  and  $\lambda \in (0, 1)$  such that  $f^0(u_2) < 0$  and  $u_1 = \lambda u_0 + (1 - \lambda)u_2$ . Then, we have

$$\begin{aligned} f^0(\lambda u_0 + (1 - \lambda)u_2) &= f^0(u_1) = 0 \quad \text{and} \\ \lambda f^0(u_0) + (1 - \lambda)f^0(u_2) &= (1 - \lambda)f^0(u_2) < 0 \\ \implies f^0(\lambda u_0 + (1 - \lambda)u_2) &\stackrel{(c)}{>} \lambda f^0(u_0) + (1 - \lambda)f^0(u_2) \end{aligned}$$

Inequality (c) contradicts the convex nature of function  $f^0$ . Thus, in turn, establishes the uniqueness of the fixed point. ■

## APPENDIX C

### INCREASING NATURE OF THE FIXED POINT

First we introduce some notation and present some preliminary observations. We will use these observations in the proof of Proposition 3.3. Let

$$h_{m,k}(q) = \sum_{k_2=0}^k \binom{k}{k_2} q^{k_2} (1 - q)^{k - k_2} P[X_{k_2} + Y_{k - k_2} \geq m]$$

where random variables  $X_{k_2}$  and  $Y_{k - k_2}$  are independent and have a binomial distribution with parameters  $(k_2, \alpha_2 u)$  and  $(k - k_2, \alpha_1 u)$ , respectively. The first step in the proof of Proposition 3.3 is to show that for any  $m \in \mathbb{Z}^+$ ,  $P[X_{k_2} + Y_{k - k_2} \geq m]$  is a non-decreasing function of  $k_2 \in [0, k]$ . Let

$$\varphi_{m,k}(k_2) = P[X_{k_2} + Y_{k - k_2} \geq m] \quad \forall k_2 \in [0, k]$$

We note that if  $u \in (0, 1]$  then  $\varphi_{m,k}(k_2) = 0$  only if  $m > k$ , i.e., if the number of active neighbours is greater than the node degree.

**Lemma C.1:** For any  $m, k \in \mathbb{Z}^+$ ,  $\varphi_{m,k}(k_2)$  is non-decreasing function of  $k_2 \in [0, k]$ .

*Proof:* Consider two integers  $k_2, k'_2 \in [0, k]$ . Without loss of generality, let  $k'_2 > k_2$  and define the following random variables

$$W_{k_2} = X_{k_2} + Y_{k - k_2} \quad \text{and} \quad W_{k'_2} = X_{k'_2} + Y_{k - k'_2}$$

Then, we have

$$\begin{aligned} W_{k_2} &=_{st} X_{k_2} + Y_{k-k_2} \\ &=_{st} \sum_{i=1}^{k_2} I_i(\alpha_2 u) + \sum_{i=1}^{k-k_2} I_i(\alpha_1 u) \end{aligned}$$

where  $I_i(p)$  are independent Bernoulli random variable with parameter  $p$  and  $=_{st}$  implies equality in distribution [26]. Therefore, for any  $k'_2 > k_2$  and  $k_2, k'_2 \in [0, k]$ , we have

$$\begin{aligned} W_{k_2} &=_{st} \sum_{i=1}^{k_2} I_i(\alpha_2 u) + \sum_{i=1}^{(k-k'_2)+(k'_2-k_2)} I_i(\alpha_1 u) \\ &=_{st} \sum_{i=1}^{k_2} I_i(\alpha_2 u) + \sum_{i=1}^{k-k'_2} I_i(\alpha_1 u) + \sum_{i=1}^{k'_2-k_2} I_i(\alpha_1 u) \\ &\stackrel{(a)}{\leq}_{st} \sum_{i=1}^{k_2} I_i(\alpha_2 u) + \sum_{i=1}^{k-k'_2} I_i(\alpha_1 u) + \sum_{i=1}^{k'_2-k_2} I_i(\alpha_2 u) \\ &=_{st} \sum_{i=1}^{k'_2} I_i(\alpha_2 u) + \sum_{i=1}^{k-k'_2} I_i(\alpha_1 u) =_{st} W_{k'_2} \end{aligned}$$

where Inequality (a) follows because  $\alpha_1 < \alpha_2 \Rightarrow \alpha_1 u \leq \alpha_2 u$ , and due to Lemma A.1. Hence, we have established that for  $k_2, k'_2 \in [0, k]$  and  $k'_2 > k_2$ , we have  $W_{k_2} \leq_{st} W_{k'_2}$ . Thus,  $\forall m \in \mathbb{Z}^+$  and  $k'_2 > k_2$ , we have

$$\begin{aligned} \varphi_{m,k}(k_2) &= P[X_{k_2} + Y_{k-k_2} \geq m] = P[W_{k_2} \geq m] \\ &\stackrel{(b)}{\leq} P[W_{k'_2} \geq m] = P[X_{k'_2} + Y_{k-k'_2} \geq m] = \varphi_{m,k}(k'_2) \end{aligned}$$

where Inequality (b) follows because  $W_{k_2} \leq_{st} W_{k'_2}$ . Thus, we have shown that  $\varphi_{m,k}(k_2)$  is a non-decreasing function of  $k_2 \in [0, k]$ . ■

Now, let us define

$$\psi_{m,k}(k_2) = \frac{k \cdot \varphi_{m,k}(k_2)}{(k - k_2)} \quad \forall k_2 \in [0, k-1]$$

Due to Lemma C.1, we can conclude that for any  $m, k \in \mathbb{Z}^+$ ,  $\psi_{m,k}(k_2)$  is also a non-decreasing function of  $k_2 \in [0, k-1]$ . Next, establish that the function  $f(q, u)$  is monotonically increasing in  $q$ .

**Proposition C.1:** If  $\alpha_2 > \alpha_1$ , then  $\frac{\partial f(q, u)}{\partial q} > 0$ .

*Proof:*

$$\begin{aligned} h_{m,k}(q) &= \sum_{k_2=0}^k \binom{k}{k_2} q^{k_2} (1-q)^{k-k_2} P[X_{k_2} + Y_{k-k_2} \geq m] \\ &= \sum_{k_2=0}^{k-1} \binom{k}{k_2} q^{k_2} (1-q)^{k-k_2} P[X_{k_2} + Y_{k-k_2} \geq m] \\ &\quad + q^m \cdot P[X_k \geq m] \\ &= (1-q) \sum_{k_2=0}^{k-1} \binom{k-1}{k_2} q^{k_2} (1-q)^{k-k_2} \cdot \psi_{m,k}(k_2) \\ &\quad + q^m \cdot P[X_k \geq m] \\ &= (1-q) \cdot \mathbb{E}[\psi_{m,k}(Z(k-1, q))] + q^m \cdot P[X_k \geq m] \end{aligned}$$

where  $Z(k-1, q)$  is a binomial random variable with parameters  $(k-1, q)$ .

Now, consider two real numbers  $q_1, q_2 \in [0, 1]$  such that

$q_1 < q_2$ . Then, we have

$$\begin{aligned} h_{m,k}(q_1) &= (1-q_1) \mathbb{E}[\psi_{m,k}(Z(k-1, q_1))] \\ &\quad + (q_1)^m P[X_k \geq m] \\ &\stackrel{(a)}{\leq} (1-q_1) \mathbb{E}[\psi_{m,k}(Z(k-1, q_1))] + (q_2)^m P[X_k \geq m] \\ &\stackrel{(b)}{\leq} (1-q_1) \mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] + (q_2)^m P[X_k \geq m] \\ &= (1-q_2) \mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] + (q_2)^m P[X_k \geq m] \\ &\quad + (q_2 - q_1) \mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] \\ &= h_{m,k}(q_2) + (q_2 - q_1) \mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] \end{aligned} \quad (11)$$

where

- (a) follows because  $q_1 < q_2$ .
- (b) since  $q_1 < q_2$ , we have  $Z(k-1, q_1) \leq_{st} Z(k-1, q_2)$  (Lemma A.1). Therefore, due to the non-decreasing nature of function  $\psi_{m,k}$ , we have  $\mathbb{E}[\psi_{m,k}(Z(k-1, q_1))] \leq \mathbb{E}[\psi_{m,k}(Z(k-1, q_2))]$  (Theorem A.1).

Also, from the definition, we also have

$$\begin{aligned} h_{m,k}(q_1) &= \mathbb{E}[\varphi_{m,k}(Z(k, q_1))] \\ &\stackrel{(c)}{\leq} \mathbb{E}[\varphi_{m,k}(Z(k, q_2))] = h_{m,k}(q_2) \end{aligned} \quad (12)$$

where Inequality (c) follows because of Lemma A.1, Lemma C.1 and Theorem A.1.

From Equations (11) and (12), we can see that for any  $q_1, q_2 \in [0, 1]$  such that  $q_1 < q_2$ , we have  $h_{m,k}(q_1) = h_{m,k}(q_2)$  if and only if  $\mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] = 0$ . From the definition of function  $\psi_{m,k}$ , we can see that  $\mathbb{E}[\psi_{m,k}(Z(k-1, q_2))] = 0$  if and only if  $m > k$ . Thus, if  $m \in [0, k]$ , then we can conclude that  $h_{m,k}(q_1) < h_{m,k}(q_2)$ .

Since the network is connected and the fraction of zealous nodes lies in the open interval  $(0, 1)$ , there exists positive integers  $k_0, m_0$  such that  $m_0 \in (0, k_0]$  and  $p_{ex}(k_0)p_{th}(m_0|k_0+1) > 0$ . Therefore, for  $q_1, q_2 \in [0, 1]$  such that  $q_1 < q_2$ , we have  $h_{m_0, k_0}(q_1) < h_{m_0, k_0}(q_2)$ . Let  $z_0 = \sum_{k \geq 1} p_{ex}(k)p_{th}(0|k+1)$  — fraction of zealous nodes. Thus, we have

$$\begin{aligned} f(q_1, u) &= z_0 + \sum_{k \geq 1} \sum_{m \geq 1} p_{ex}(k)p_{th}(m|k+1)h_{m,k}(q_1) \\ &= \sum_{k \geq 1} \sum_{m \geq 1} p_{ex}(k)p_{th}(m|k+1)h_{m,k}(q_1)(1 - \mathbb{I}_{\left\{ \substack{m=m_0 \\ k=k_0} \right\}}) \\ &\quad + p_{ex}(k_0)p_{th}(m_0|k_0+1) \cdot h_{m_0, k_0}(q_1) + z_0 \\ &< \sum_{k \geq 1} \sum_{m \geq 1} p_{ex}(k)p_{th}(m|k+1)h_{m,k}(q_1)(1 - \mathbb{I}_{\left\{ \substack{m=m_0 \\ k=k_0} \right\}}) \\ &\quad + p_{ex}(k_0)p_{th}(m_0|k_0+1) \cdot h_{m_0, k_0}(q_2) + z_0 \\ &\leq \sum_{k \geq 1} \sum_{m \geq 1} p_{ex}(k)p_{th}(m|k+1)h_{m,k}(q_2)(1 - \mathbb{I}_{\left\{ \substack{m=m_0 \\ k=k_0} \right\}}) \\ &\quad + p_{ex}(k_0)p_{th}(m_0|k_0+1) \cdot h_{m_0, k_0}(q_2) + z_0 \\ &= f(q_2, u) \end{aligned}$$

i.e., for  $q_1, q_2 \in [0, 1]$  such that  $q_1 < q_2$ , we have  $f(q_1, u) < f(q_2, u)$ . ■

### A. Proof of Proposition 3.3

The continuity of the fixed point follows from the fact that function  $f(\cdot, u)$  is a continuous map for all  $u \in [0, 1]$ . To prove the increasing nature of the fixed point, we use the *implicit function theorem*.

Let  $f^0(q, u) = f(q, u) - u$ . According to the implicit function theorem, when  $f^0(q, u) = 0$ , we have

$$\frac{du}{dq} = -\frac{\frac{\partial f^0(q, u)}{\partial q}}{\frac{\partial f^0(q, u)}{\partial u}} = \frac{\frac{\partial f(q, u)}{\partial q}}{\left(1 - \frac{\partial f(q, u)}{\partial u}\right)}$$

From Proposition 3.2, we know that for every  $q \in [0, 1]$ , there is a unique fixed point  $u_q \in (0, 1)$  that satisfies  $f(q, u_q) = u_q$  i.e.,  $f^0(q, u_q) = 0$ . Therefore, we have

$$\frac{du_q}{dq} = \frac{\frac{\partial f(q, u_q)}{\partial q}}{\left(1 - \frac{\partial f(q, u)}{\partial u}\right)\Big|_{u=u_q}}$$

From Proposition C.1, we know that  $\frac{\partial f(q, u)}{\partial q} > 0, \forall u \in (0, 1)$ . Since  $u_q \in (0, 1)$ , we have  $\frac{\partial f(q, u_q)}{\partial q} > 0$ . Now, to prove Proposition 3.3, we just need to show that  $\frac{\partial f(q, u)}{\partial u}\Big|_{u=u_q} < 1$ .

We show this by contradiction. We note that  $f(q, 1) < 1$ , where Inequality (a) follows because the  $f(q, 1)$  is a convex combination of non-negative terms less than or equal to 1 (some terms are strictly less than 1). Since the function  $f(q, u)$  is continuous and non-decreasing in  $u$  (see Proposition 3.1), there exists a point  $u_q \in (u_q, 1)$  such that  $f(q, u_q) < u_q$ .

We note that the function  $f(q, u)$  is convex in  $u \in (0, 1)$  (see Proposition 3.1). Now, let us assume that  $\frac{\partial f(q, u)}{\partial u}\Big|_{u=u_q} \geq 1$ . Then, from the definition of convex functions, we have

$$\begin{aligned} f(q, u'_q) &\geq f(q, u_q) + (u'_q - u_q) \cdot \frac{\partial f(q, u)}{\partial u}\Big|_{u=u_q} \\ &\geq f(q, u_q) + (u'_q - u_q) \\ &\stackrel{(b)}{=} u_q + (u'_q - u_q) = u'_q \\ &\Rightarrow f(q, u'_q) \stackrel{(c)}{\geq} u'_q \end{aligned}$$

where Inequality (b) follows because we have  $f(q, u_q) = u_q$ .

Inequality (c) contradicts the fact that  $f(q, u_q) < u_q$ . Since this contradiction was due to the assumption that  $\frac{\partial f(q, u)}{\partial u}\Big|_{u=u_q} \geq 1$ , we should have  $\frac{\partial f(q, u)}{\partial u}\Big|_{u=u_q} < 1$ . This, in turn, establishes that  $\frac{du_q}{dq} > 0$ . ■

## APPENDIX D

### EQUIVALENCE TO THE LINEAR PROGRAM

In this section, we prove the equivalence between problems  $P_1$  and  $P_2$  presented in Section IV

### A. Proof of Proposition 4.3

Let the value of the objective of problem  $P_1$  for incentive-policy  $\phi$  be denoted as  $c(\phi)$ . Let  $\phi^o = \{\phi_k^o, k \geq 1\}$  be an optimal solution of problem  $P_1$ . If we have  $\sum_{k \geq 1} k \cdot p(k) \cdot \phi^o(k) = \bar{d} \cdot q_\gamma$ , then we can choose  $\phi^{opt} = \phi^o$ .

So, let us assume that  $\frac{1}{\bar{d}} \sum_{k \geq 1} k \cdot p(k) \cdot \phi^o(k) = q^o > q_\gamma$ . Our goal is to obtain a policy  $\phi^{opt}$  from  $\phi^o$  without increasing the cost. Let  $\phi^{opt} = \frac{q_\gamma}{q^o} \phi^o$ . Then, we have

$$\begin{aligned} \frac{1}{\bar{d}} \sum_{k \geq 1} k p(k) \cdot \phi^{opt}(k) &= \frac{q_\gamma}{q^o \bar{d}} \sum_{k \geq 1} k \cdot p(k) \cdot \phi^o(k) = q_\gamma \quad (13) \\ \text{and } 0 &\leq \phi^{opt} \leq \phi^o \leq 1 \quad (14) \end{aligned}$$

From Equality (13) and Inequality (14) we can see that  $\phi^{opt}$  is a feasible solution of problem  $P_1$ . Further, we also have

$$\begin{aligned} c(\phi^o) &= \sum_{k \geq 1} c_k \cdot p(k) \cdot \phi^o(k) \cdot s_k(q^o) \\ &\stackrel{(a)}{\geq} \sum_{k \geq 1} c_k \cdot p(k) \cdot \phi^{opt}(k) \cdot s_k(q^o) \\ &\stackrel{(b)}{\geq} \sum_{k \geq 1} c_k \cdot p(k) \cdot \phi^{opt}(k) \cdot s_k(q_\gamma) \\ &= c(\phi^{opt}) \stackrel{(c)}{\geq} c(\phi^o) \Rightarrow c(\phi^o) = c(\phi^{opt}) \end{aligned}$$

where Inequality (a) follows because  $\phi^{opt} \leq \phi^o$ , Inequality (b) follows because  $q^o > q_\gamma$  and  $s_k(q)$  is non-decreasing in  $q$  (Proposition 4.2), and Inequality (c) follows because  $\phi^o$  is an optimal solution of problem  $P_1$ .

Since the cost for policies  $\phi^o$  and  $\phi^{opt}$  are equal, and  $\phi^o$  is an optimal policy, we can conclude that policy  $\phi^{opt}$  is also an optimal solution of problem  $P_1$ . ■

### B. Proof of Proposition 4.4

Let  $\nu^{opt} = \{\nu_k^{opt}, k \geq 1\}$  be an optimal solution of problem  $P_2$ . Let us define

$$\phi^f(k) = \begin{cases} 0 & \text{if } p(k) = 0 \\ \frac{\nu_k^{opt} \cdot \bar{d} \cdot q_\gamma}{k \cdot p(k)} & \text{otherwise} \end{cases}$$

Then, we have

$$\frac{1}{\bar{d}} \sum_{k \geq 1} k \cdot p(k) \cdot \phi^f(k) = q_\gamma \sum_{k \geq 1} \nu_k^{opt} \stackrel{(a)}{=} q_\gamma \quad (15)$$

$$\phi^f(k) = \begin{cases} 0 < 1 & \text{if } p(k) = 0 \\ \frac{\nu_k^{opt} \cdot \bar{d} \cdot q_\gamma}{k \cdot p(k)} \leq 1 & \text{otherwise} \end{cases} \quad (16)$$

where Equality (a) and Inequality (b) follows because  $\{\nu_k^{opt}, k \geq 1\}$  is a feasible solution of problem  $P_2$ . From Equality (15) and Inequality (16), we can conclude that  $\phi^f$  is a feasible solution of problem  $P_1$ .

Now, let  $\phi^{opt} = \{\phi_k^{opt}, k \geq 1\}$  be an optimal solution of problem  $P_1$  such that  $\sum_{k \geq 1} k \cdot p(k) \cdot \phi^{opt}(k) = \bar{d} \cdot q_\gamma$ . The existence of such an optimal policy is guaranteed by Proposition 4.3. Let us define

$$\nu^f(k) = \frac{k \cdot p(k) \cdot \phi_k^{opt}}{\bar{d} \cdot q_\gamma}$$

Then, we have  $\sum_{k \geq 1} \nu^f(k) = \sum_{k \geq 1} \frac{k \cdot p(k) \cdot \phi_k^{opt}}{\bar{d} \cdot q_\gamma} = 1$  and  $0 \leq \nu_k = \frac{k \cdot p(k) \cdot \phi_k^{opt}}{\bar{d} \cdot q_\gamma} \leq \frac{k \cdot p(k)}{\bar{d} \cdot q_\gamma}$ . Hence,  $\nu^f$  is a feasible solution of

problem  $P_2$ . Further, we also have

$$\begin{aligned}
& \sum_{k \geq 1} c_k p(k) \phi^{opt}(k) s_k(q_\gamma) \leq \sum_{k \geq 1} c_k p(k) \phi^f(k) s_k(q_\gamma) \\
& = \bar{d} \cdot q_\gamma \sum_{k \geq 1} \nu_k^{opt} \cdot \mu_k(q_\gamma) \leq \bar{d} \cdot q_\gamma \sum_{k \geq 1} \nu_k^f \cdot \mu_k(q_\gamma) \\
& = \sum_{k \geq 1} c_k p(k) \phi^{opt}(k) s_k(q_\gamma) \\
& \Rightarrow \sum_{k \geq 1} c_k p(k) \phi^{opt}(k) s_k(q_\gamma) = \sum_{k \geq 1} c_k p(k) \phi^f(k) s_k(q_\gamma) \\
& \Rightarrow \phi^f \text{ is an optimal solution of problem } P_1.
\end{aligned}$$

## APPENDIX E

### OPTIMALITY OF ALGORITHM 2

In this section, we prove that the vector  $\nu^a$  returned by Algorithm 2 is an optimal point of problem  $P_2$ . However, as a first step, we prove the following lemma. Let  $\mathcal{S} = \{1, 2, \dots, k_{max}\}$  and  $k_{max}$  is the maximum degree in the network.

**Lemma E.1:** Let  $\nu$ , a  $1 \times k_{max}$  vector, be a feasible solution of problem  $P_2$ . Then, we have

$$\sum_{j=1}^{j_0} \nu_{\sigma(j)}^a \geq \sum_{j=1}^{j_0} \nu_{\sigma(j)} \quad \forall j_0 \in \mathcal{S}$$

*Proof:* We prove this lemma by induction. For  $j_0 = 1$ , we have

$$\nu_{\sigma(1)}^a = \min \left\{ 1, \frac{\sigma(1) \cdot p(\sigma(1))}{\bar{d} \cdot q_\gamma} \right\} \stackrel{(a)}{\geq} \nu_{\sigma(1)} \quad (17)$$

where Inequality (a) follows because  $\nu$  is a feasible solution of problem  $P_2$ . Now, let us assume that  $\sum_{j=1}^{j_1} \nu_{\sigma(j)}^a \geq \sum_{j=1}^{j_1} \nu_{\sigma(j)}$  for some  $j_1 \in \mathcal{S}$ . We need to show that  $\sum_{j=1}^{j_1+1} \nu_{\sigma(j)}^a \geq \sum_{j=1}^{j_1+1} \nu_{\sigma(j)}$ .

If we have  $\nu_{\sigma(j_1+1)}^a \geq \nu_{\sigma(j_1+1)}$ , then the result is trivial. Let us therefore consider the case where  $\nu_{\sigma(j_1+1)}^a < \nu_{\sigma(j_1+1)}$ . This implies that

$$\begin{aligned}
\nu_{\sigma(j_1+1)}^a < \nu_{\sigma(j_1+1)} &\leq \frac{\sigma(j_1+1) \cdot p(\sigma(j_1+1))}{\bar{d} \cdot q_\gamma} \\
\Rightarrow \nu_{\sigma(j_1+1)}^a &\stackrel{(b)}{=} 1 - \sum_{j=1}^{j_1} \nu_{\sigma(j)}^a
\end{aligned}$$

where Equality (b) follows because of steps 4 and 5 of Algorithm 2. Therefore, if  $\nu_{\sigma(j_1+1)}^a < \nu_{\sigma(j_1+1)}$ , then we have

$$\sum_{j=1}^{j_1+1} \nu_{\sigma(j)}^a = 1 = \sum_{j \in \mathcal{S}} \nu_{\sigma(j)} \geq \sum_{j=1}^{j_1+1} \nu_{\sigma(j)}$$

#### A. Proof of Proposition 4.5

Let  $\nu$  be a feasible solution of problem  $P_2$ . Since  $\nu^a$  and  $\nu$  are feasible solutions of problem  $P_2$ , we have  $\sum_{j \geq 1} \nu^a(\sigma(j)) = \sum_{j \geq 1} \nu(\sigma(j)) = 1$ . Let  $X^a$  and  $X$  two independent random variable taking values from the set  $\mathcal{S}$  such that  $P[X^a = j] = \nu_{\sigma(j)}^a$  and  $P[X = j] = \nu_{\sigma(j)}$ . Then, from Lemma E.1, we have

$$P[X^a \leq j_0] = \sum_{j=1}^{j_0} \nu_{\sigma(j)}^a \geq \sum_{j=1}^{j_0} \nu_{\sigma(j)} = P[X \leq j_0] \quad (18)$$

From Equation 18 and the definition of usual stochastic orders we can see that  $X^a \leq_{st} X$ . Therefore, we have

$$\begin{aligned}
& \sum_{k \geq 1} \nu_k^a \cdot \mu_k(q_\gamma) = \sum_{j \in \mathcal{S}} \nu_{\sigma(j)}^a \cdot \mu_{\sigma(j)}(q_\gamma) \\
& = \sum_{j \in \mathcal{S}} P[X^a = j] \cdot \mu_{\sigma(j)}(q_\gamma) = \mathbb{E}[\mu_{\sigma(X^a)}(q_\gamma)] \\
& \stackrel{(a)}{\leq} \mathbb{E}[\mu_{\sigma(X)}(q_\gamma)] = \sum_{j \in \mathcal{S}} P[X = j] \cdot \mu_{\sigma(j)}(q_\gamma) \\
& = \sum_{j \in \mathcal{S}} \nu_{\sigma(j)} \cdot \mu_{\sigma(j)}(q_\gamma) = \sum_{k \geq 1} \nu_k \cdot \mu_k(q_\gamma) \quad (19)
\end{aligned}$$

where Inequality (a) follows from an application of Theorem A.1 ( $X^a \leq_{st} X$  and  $\mu_{\sigma(j)}(q_\gamma)$  is a non-decreasing function of  $j$ ).

Let  $\nu^{opt}$  be an optimal solution of  $P_2$ . Then, we have the following inequalities

$$\begin{aligned}
& \sum_{k \geq 1} \nu_k^{opt} \cdot \mu_k(q_\gamma) \leq \sum_{k \geq 1} \nu_k^a \cdot \mu_k(q_\gamma) \stackrel{(b)}{\leq} \sum_{k \geq 1} \nu_k^{opt} \cdot \mu_k(q_\gamma) \\
& \Rightarrow \sum_{k \geq 1} \nu_k^{opt} \cdot \mu_k(q_\gamma) = \sum_{k \geq 1} \nu_k^a \cdot \mu_k(q_\gamma) \\
& \Rightarrow \nu^a \text{ is an optimal solution of } P_2
\end{aligned}$$

where Inequality (b) follows by setting  $\nu$  to  $\nu^{opt}$  in Inequality (19). ■

## REFERENCES

- [1] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. 15th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2009, pp. 497–506.
- [2] J. Leskovec, A. Singh, and J. Kleinberg, "Patterns of influence in a recommendation network," in *Advances in Knowledge Discovery and Data Mining*. Springer, 2006, pp. 380–389.
- [3] T. N. Dinh, D. T. Nguyen, and M. T. Thai, "Cheap, easy, and massively effective viral marketing in social networks: truth or fiction?" in *Proc. 23rd ACM Conf. Hypertext and Social Media*, 2012, pp. 165–174.
- [4] D. Houston, "Dropbox - startup lessons learned," *Presentation Slides available at <http://www.slideshare.net/guest94e4c/dropbox-startup-lessons-learned-3836587>*, 2010.
- [5] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2003, pp. 137–146.
- [6] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2009, pp. 199–208.
- [7] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. 10th IEEE Int. Conf. Data Mining (ICDM)*, 2010, pp. 88–97.
- [8] J. Hartline, V. Mirrokni, and M. Sundararajan, "Optimal marketing strategies over social networks," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 189–198.
- [9] D. Arthur, R. Motwani, A. Sharma, and Y. Xu, "Pricing strategies for viral marketing on social networks," in *Internet and Network Economics*. Springer, 2009, pp. 101–112.
- [10] I. Lobel, E. D. Sadler, and L. R. Varshney, "Customer referral incentives and social media," *Available at SSRN: <http://ssrn.com/abstract=2520615>*, 2014.
- [11] A. Karnik and P. Dayama, "Optimal control of information epidemics," in *Proc. 4th IEEE Int. Conf. Communication Systems and Networks (COMSNETS)*, 2012, pp. 1–7.
- [12] P. Dayama, A. Karnik, and Y. Narahari, "Optimal incentive timing strategies for product marketing on social networks," in *Proc 11th Int. Conf. Autonomous Agents and Multiagent Systems-Volume 2*, 2012, pp. 703–710.
- [13] K. Kandhway and J. Kuri, "How to run a campaign: Optimal control of sis and sir information epidemics," *Applied Mathematics and Computation*, vol. 231, no. 0, pp. 79 – 92, 2014.
- [14] —, "Optimal control of information epidemics modeled as maki thompson rumors," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 12, pp. 4135–4147, 2014.

- [15] —, “Campaigning in heterogeneous social networks: Optimal control of si information epidemics,” *IEEE/ACM Trans. Netw.*, vol. PP, no. 99, pp. 1–1, 2014.
- [16] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical processes on complex networks*. Cambridge University Press, 2008.
- [17] G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes, “Bootstrap percolation on complex networks,” *Phys. Rev. E*, vol. 82, no. 1, p. 011103, 2010.
- [18] S. N. Dorogovtsev, *Lectures on complex networks*. Oxford University Press Oxford, 2010, vol. 24.
- [19] M. Granovetter, “Threshold models of collective behavior,” *American journal of sociology*, pp. 1420–1443, 1978.
- [20] M. W. Macy, “Chains of cooperation: Threshold effects in collective action,” *American Sociological Review*, pp. 730–747, 1991.
- [21] V. Mahajan, E. Muller, and F. M. Bass, “New product diffusion models in marketing: A review and directions for research,” in *Diffusion of technologies and social behavior*. Springer, 1991, pp. 125–177.
- [22] E. Berger, “Dynamic monopolies of constant size,” *Journal of Combinatorial Theory, Series B*, vol. 83, no. 2, pp. 191–200, 2001.
- [23] N. Karmarkar, “A New Polynomial-time Algorithm for Linear Programming,” in *Proc. 16th Annu. ACM Symp. Theory of Computing*, 1984, pp. 302–311.
- [24] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” <http://snap.stanford.edu/data>, Jun. 2014.
- [25] J. Kunegis, “KONECT Datasets: Koblenz network collection,” <http://konect.uni-koblenz.de/>, 2015.
- [26] M. Shaked and J. G. Shanthikumar, *Stochastic Orders*. Springer, 2007.